

TUMSAT-OACIS Repository - Tokyo University of Marine Science and Technology (東京海洋大学)

鯨体計測データを用いた属性判別解析に対する統計的学習法の適用とその精度評価に関する研究

著者	高橋 萌
学位名	博士（海洋科学）
学位授与機関	東京海洋大学
学位授与年度	2017
学位授与番号	12614博甲第506号
権利	全文公表年月日：2019-06-24
URL	http://id.nii.ac.jp/1342/00001726/

博士學位論文内容要旨
Abstract

専攻 Major	応用環境システム学	氏名 Name	高橋 萌
論文題目 Title	鯨体計測データを用いた属性判別解析に対する統計的学習法の適用と その精度評価に関する研究		

鯨類資源の保護・管理において、系群をはじめとする集団や性などの属性の判別・予測は重要な要素となる。鯨類の外部・頭骨形態は、生息環境や生活史、それに伴う摂餌生態などの影響を受けるため、個体の属性に関する様々な情報が含まれる。従って、形態学的情報は、集団の判別や集団構造の推定に関して、遺伝学や生態学などの知見と並んで有益な情報源とされている。そのため、それらの形態的特徴を適切に把握できれば様々な属性の判別に寄与することが期待される。しかし鯨類では、混獲やストランディングなど生体標本の採集機会が少なく、また捕獲採取にかかる労力も大きいことから標本数が小さくなる傾向にある。また、性別などの属性が不明な場合や、標本数の偏りを伴う場合も多いことなどから、適用できる判別手法や判別・予測の精度には限界があった。そこで本研究では、従来の判別解析に加え、統計的学習法も活用し、鯨類の形態計測データを用いた属性判別解析における種々の課題の克服や精度向上が可能なかを、シミュレーション実験を通して統計的に評価するとともに、実データへの適用を通してその有用性と実用性について検討した。

第1章を本稿の導入部とし、鯨類資源解析における形態測定データの役割や、統計的学習法を用いた属性判別の理論的背景、ならびにランダムフォレスト (RF) など本稿で扱った手法と判別器の評価方法について述べた。

第2章では、雌雄各6個体のシャチ *Orcinus orca* の頭骨全18部位の形態測定データによる雌雄判別器の構築を試み、標本数に対し計測項目が高次元となるような小標本時における統計的学習法の有用性を検討した。シミュレーションの結果、従来型の判別解析では、標本数に対し計測項目の多いデータに対応するため、変数選択を伴う必要が生じ、元のデータに適合しすぎる過学習な判別規則を提示する可能性が示唆された。一方、機械学習法の一つである RF では過学習となる可能性が低く、小標本の場合であっても高次元情報を生かした判別規則の導出が可能であると示された。

第3章では、標本数やその偏りが予測精度に及ぼす影響について検討した。RF では小標本下での精度に弱点は見られたものの、標本数の増加に伴い確実に精度が向上することや、過学習が生じないことが利点として明らかになった。この結果は、既に他分野における先行研究でも報告されていたが、鯨体計測データのような小標本かつ変数過多となりやすい条件での検証は、科学的に重要な知見となる。一方 LDA では、一定の標本数に達した時点で RF の予測精度に及ばなくなったものの、対象としたデータでは、標本数の偏りに対して影響を受けにくいことが明らかとなった。

第4章では、第一期/二期北西太平洋鯨類捕獲調査 (1994-2010 年) で捕獲された北西太平洋産ミンククジラ *Balaenoptera acutorostrata* の2集団、合計915個体の性成熟個体を対象とし、系群判別器の構築を試みた。遺伝により推定された系群情報を教師ラベルとし、頭骨と外部形態の全20部位の計測データを基にした系群判別器の実行可能性と有用性について、①判別手法と特徴量の選択による精度への影響、②教師個体の選出パターンによる精度への影響、③標本数の偏り補正による判別精度への効果、の3点を通して検討した。判別手法ごとに、特徴量の変換方法や量の多少による精度への影響は異なり、それらの最適な組み合わせが精度向上に必要であった。また、判別の目的に応じた評価指標の設定と、複数の指標による多面的な評価の必要性が示唆された。さらに、同一集団であっても、教師に選出する個体の違いは判別精度に影響し、生物学的な情報を考慮した教師個体の選別・層化の検討、あるいは形態計測データに付随する年齢や肉体的成熟の指標が必要となる可能性が示唆された。標本数の偏り補正にはダウンサンプリング法を検討したが、標本数が少ない集団の判別精度の向上と引き換えに、標本数が多い集団の再現率が低下するというトレードオフの関係を示した。一方、RF のアルゴリズム内での改善により精度向上が認められ、判別性能の向上には、属性の特徴をより強く反映した形態的特徴量の探索か、情報量を減らさずに効率的に学習できるようアルゴリズム内での改善が必要であると考えられた。

第5章では、各章の検討で得られた統計的学習法の導入に関する知見を取りまとめ、形態計測データを用いた属性判別解析への有用性について考察した。限られた標本数と計測部位から得られる形態

学的情報には限界があり，いかに有効利用できるかは，形態計測データに対して適切な前処理や判別手法を適用できるかに依存していた．それゆえに，従来の判別解析に加えて，それぞれの弱点と利点を考慮した統計的学習法の利用は，鯨類の形態情報に基づく判別の可能性を広げ，潜在的な価値の向上につながると考えられた．また，形態情報を主軸とした属性予測の更なる精度向上には，判別手法あるいは手続きの改善の他，形態情報の多様化や補助的な情報を追加することが望まれ，今後の課題とした．
